Providing Predictive Models for Hazard Identification

Gilman D. Veith

2003 EPA Computational Toxicology Workshop

September 29-30, 2003

Framework Includes Three Major Topics under Models for Hazard Identification

- QSAR and Other Computational Approaches
  - Early OPPTS applications in New Chemical Review
- Pollution Prevention Strategies
  - OPPTS leadership with web-based PBT Profilers
- High Through-Put Screening
  - Opportunity to integrate with other required testing

QSAR-Based Models Can Serve Three Roles in Risk Assessment

- Missing Data for Untested Chemicals
  - -frequent domain violations for complex endpoints
  - -regulatory acceptancy criteria for QSAR by EU
- Prioritization for Chemical Assessments
  - -broad classification of chemical attributes (PBT)
  - -relative ranking of specific hazards
- Hypothesis-Driven Strategic Testing
  - -minimize the generation of unused test data
  - -optimize the selection of dose-response tests

Marking Progress in QSAR

"QSAR may predict chemical properties, but it will never predict bioaccumulation and residues"---ASTM workshop…1973.

"QSAR may be able to predict bioaccumulation, but it will never predict toxicity of chemicals"---EPA Bioassay Ctte…1979

"OK, QSAR may be able to predict acute toxicity, but it will never be able to predict chronic toxicity"---Bioassay Ctte…1983

"QSAR didn't work for health effects in the 60's and it won't work now" ---QSAR budget cut justification, ORD AA…1984

"QSAR may work for simple chemicals, but it will never predict complex phenomena like metabolism"---QSAR workshop…1992

"QSAR may work for ecotoxicology, but it will never work for chronic, non-cancer endpoints"---EPA risk manager…1999

Complexity in Quantifying Chemical Structure

- **Conformational Analysis**

  Chemical properties/interactions are highly sensitive to induced conformational changes

  - which structure should be quantified?

- Metabolic Simulation

  Many toxic effects do not arise from nor can they be forecasted directly from the parent chemical

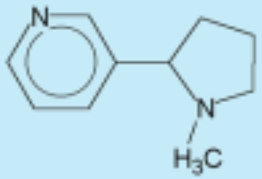
  – which metabolite should be modeled?

- Chemical Inventories and Scaffolds

  Environmental risks include those from the few tested chemicals and thousands of untested chemicals

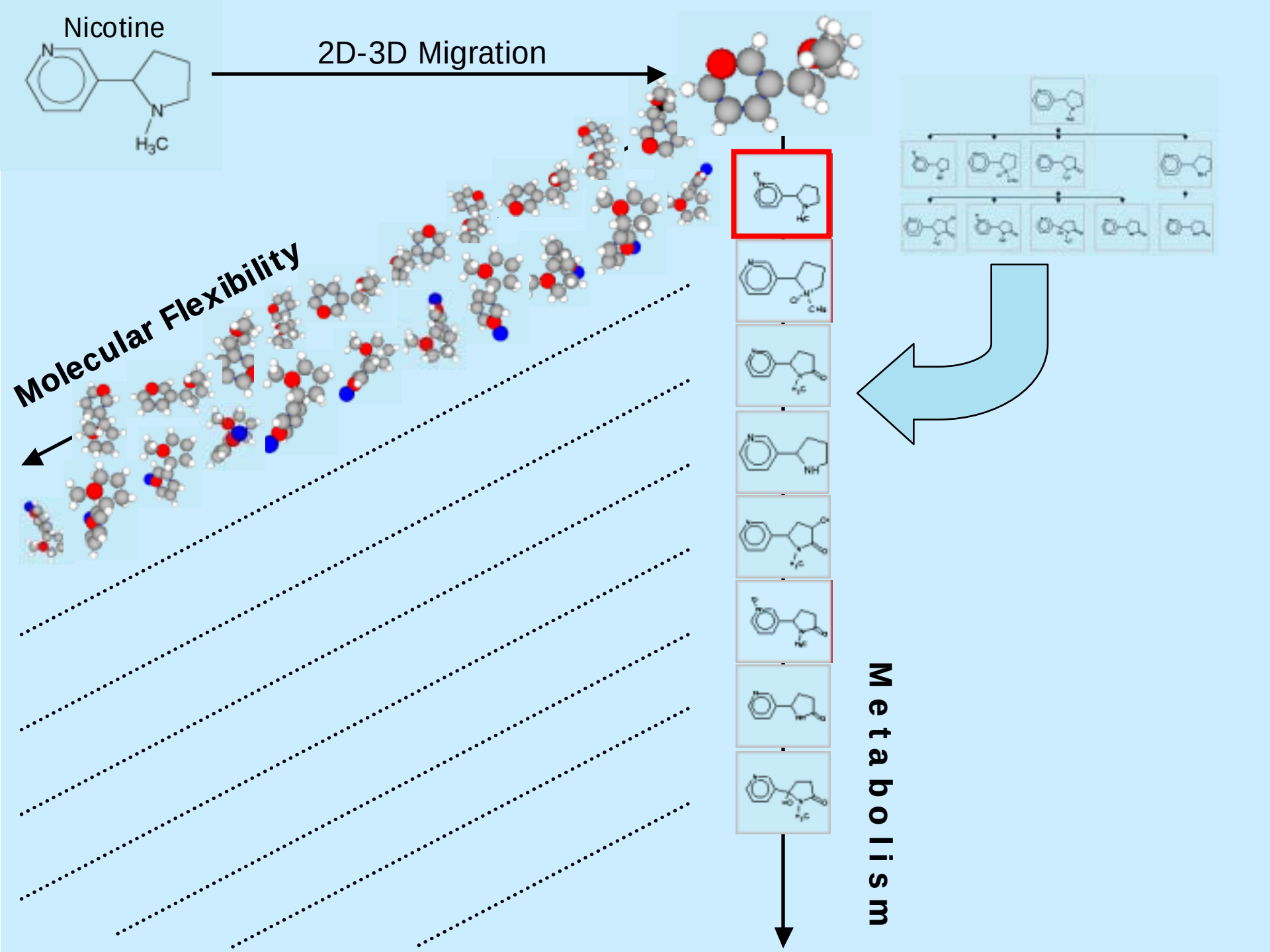
  - which group/list should be EPA's priority?

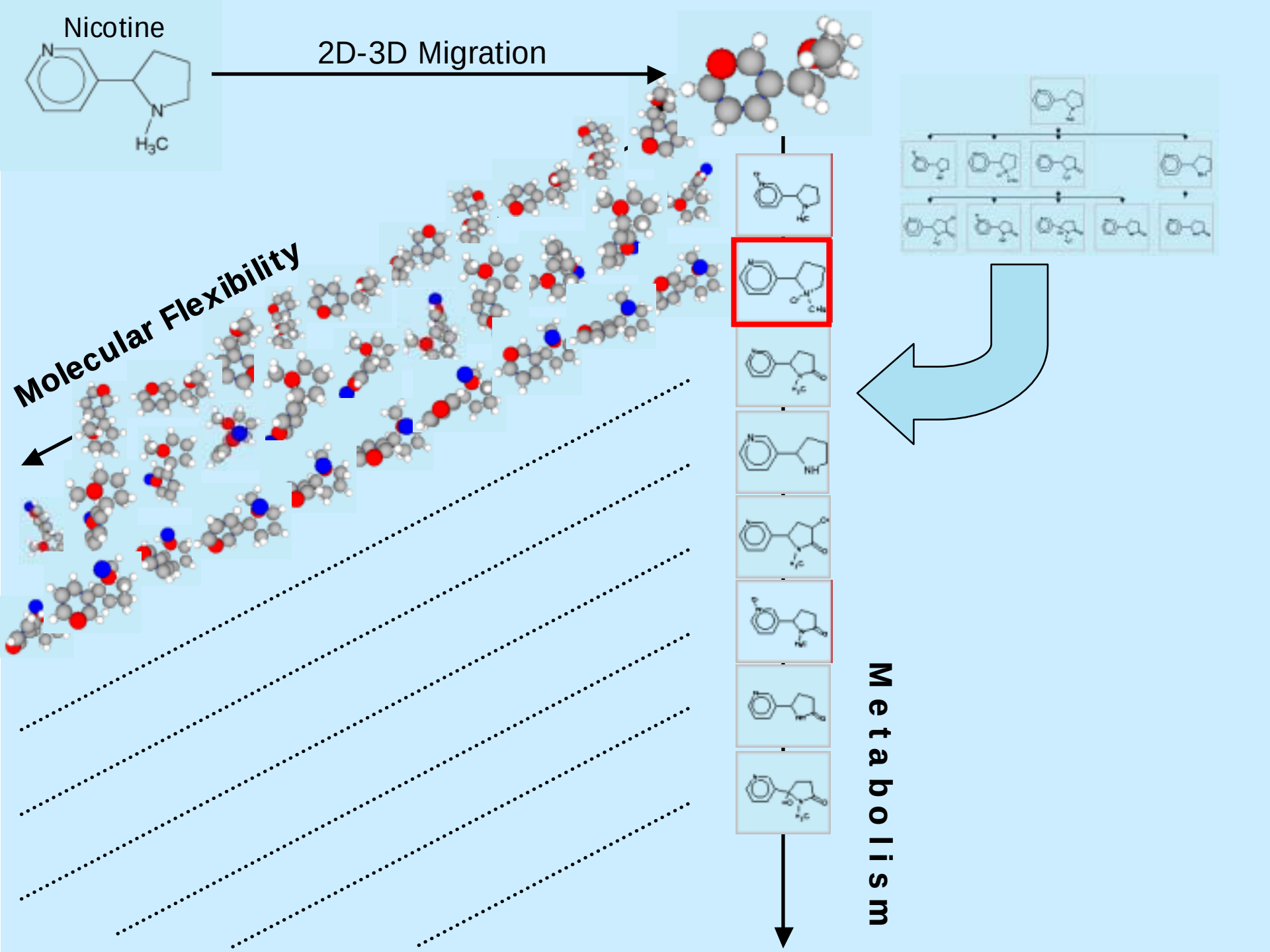
Nicotine

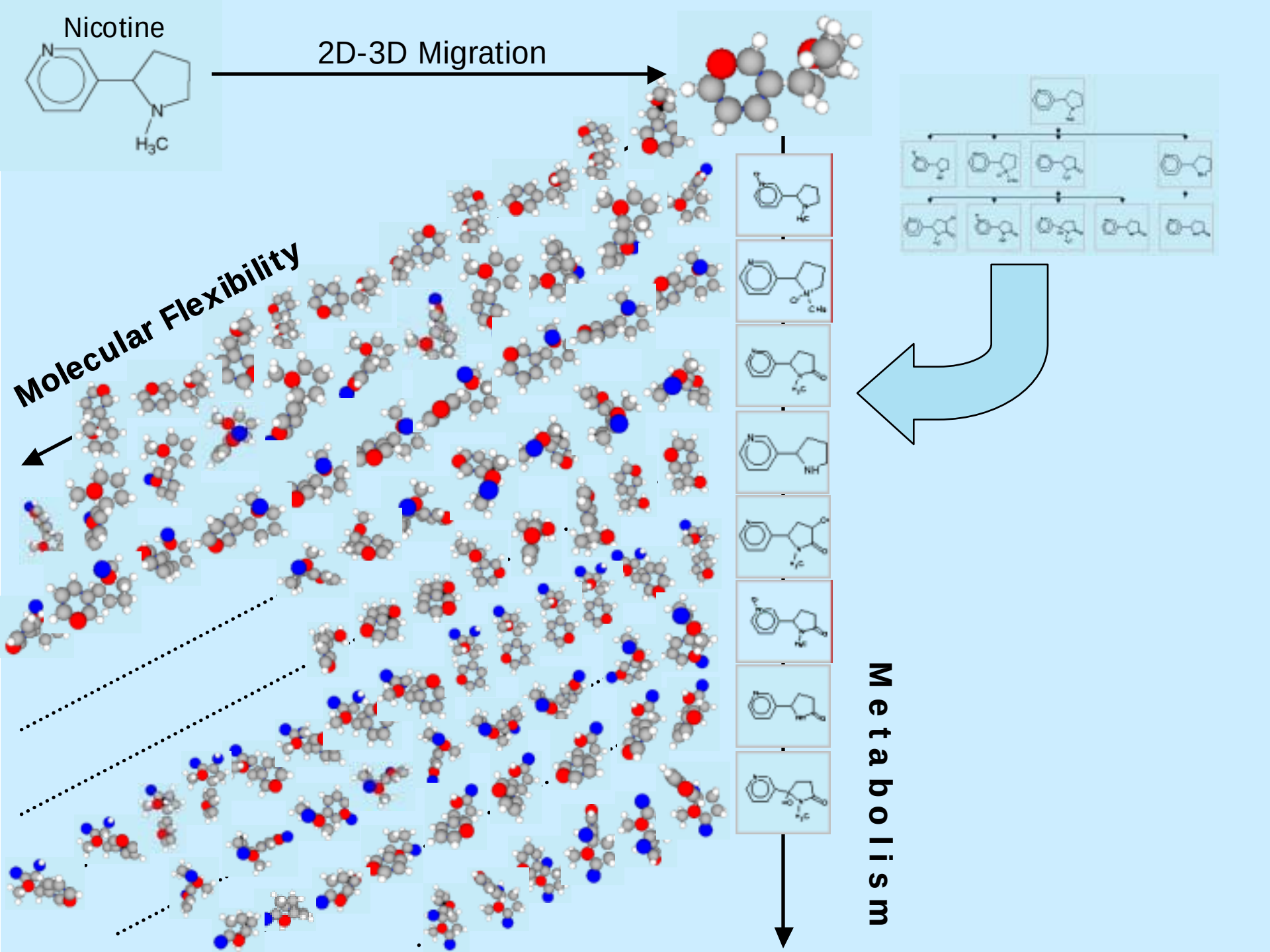


2D-3D Migration

Nicotine
2D-3D Migration
Molecular Flexibility
Rat Liver Metabolism
Metabolism


Nicotine
2D-3D Migration
Molecular Flexibility
Metabolism


Nicotine
2D-3D Migration
Molecular Flexibility
Metabolism


Nicotine
2D-3D Migration
Molecular Flexibility
Metabolism

Complexity in Quantifying Chemical Structure

- Conformational Analysis

  Chemical properties/interactions are highly sensitive to induced conformational changes

  - which structure should be quantified?

- **Metabolic Simulation**

  Many toxic effects do not arise from nor can they be forecasted directly from the parent chemical

  - which metabolite should be modeled?

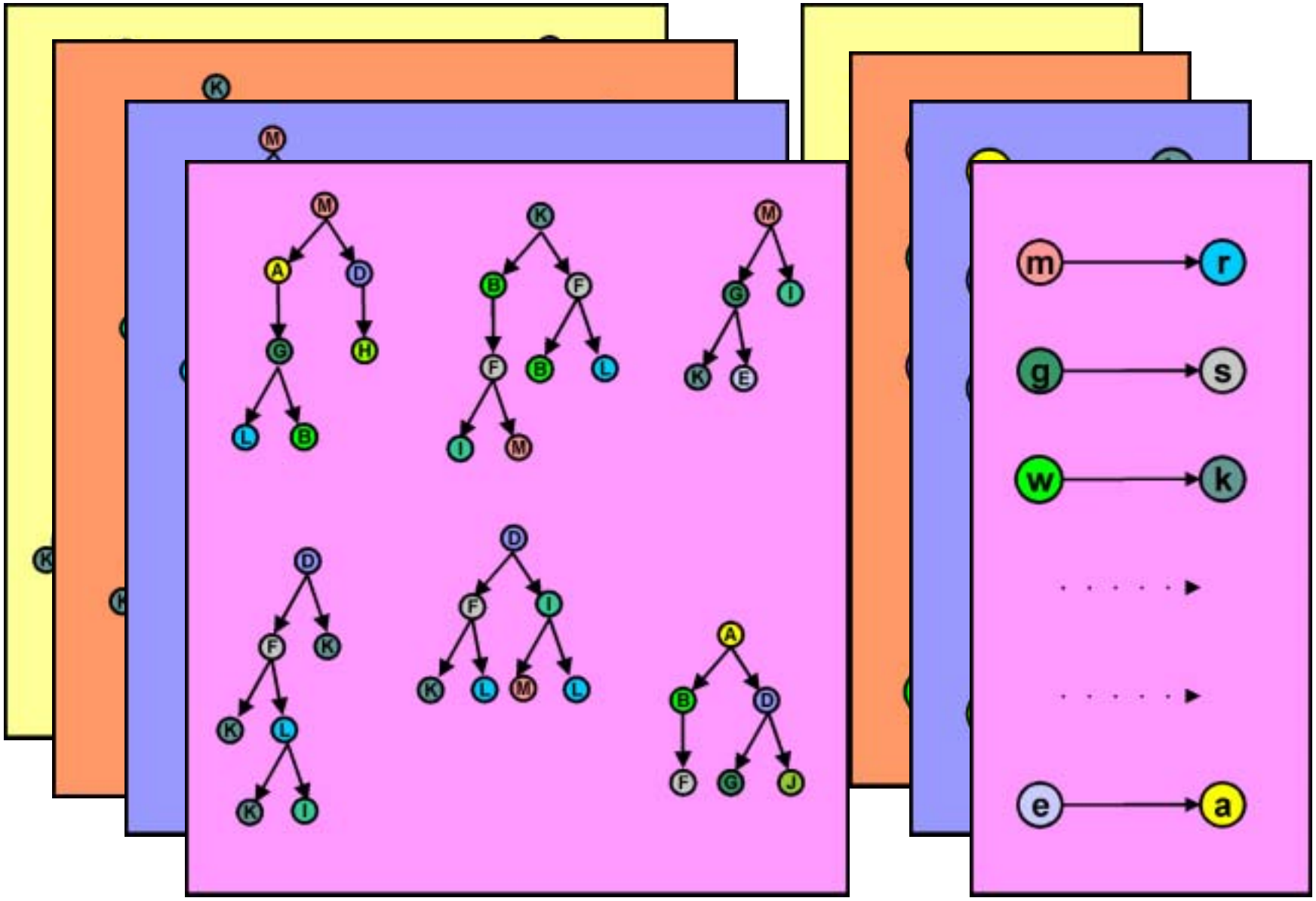
- Chemical Inventories and Scaffolds

  Environmental risks include those from the few tested chemicals and thousands of untested chemicals

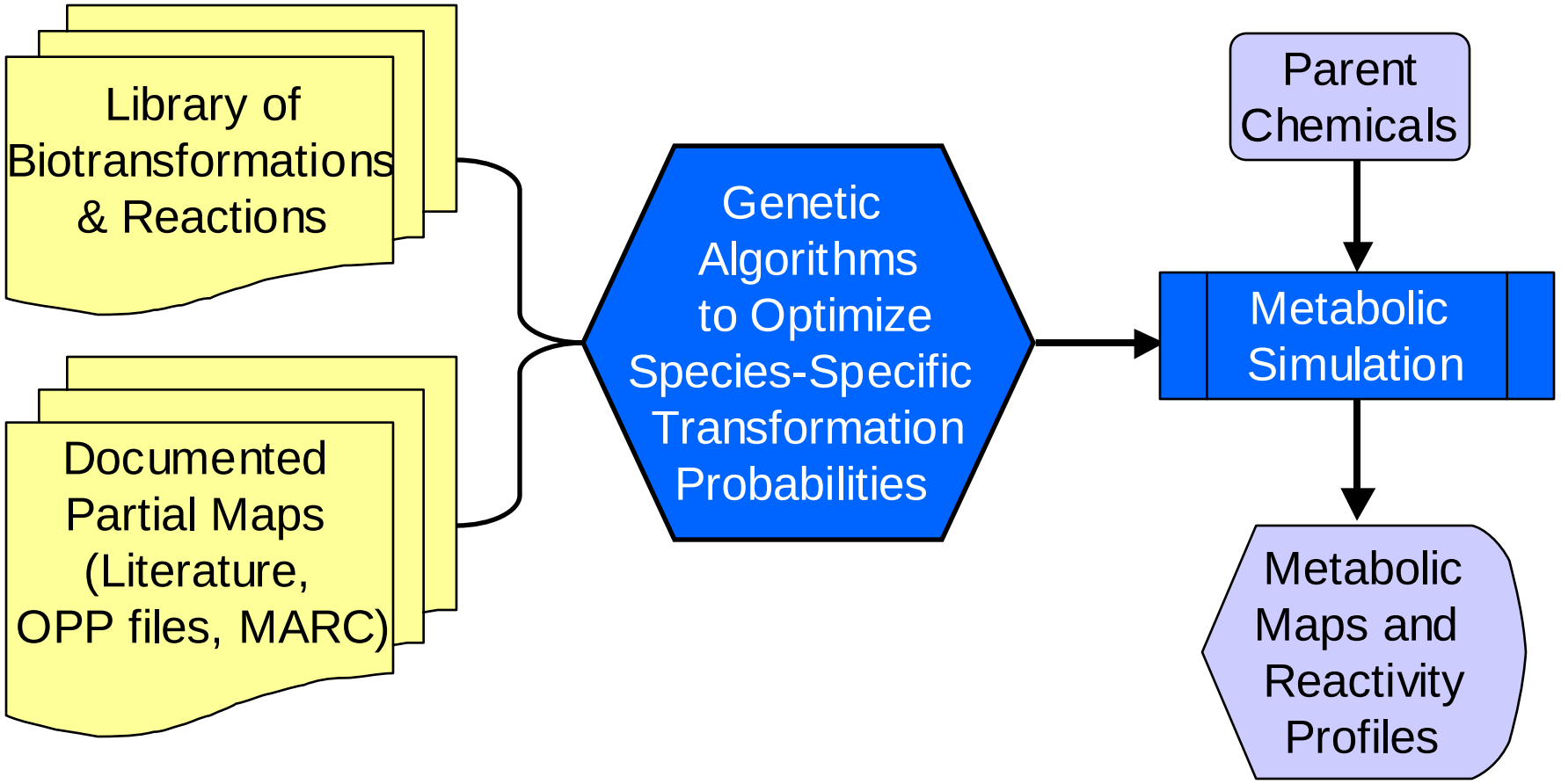
  - which group/list should be EPA's priority?

Documented maps
tissue/organ specific

List of transformations
tissue/organ specific

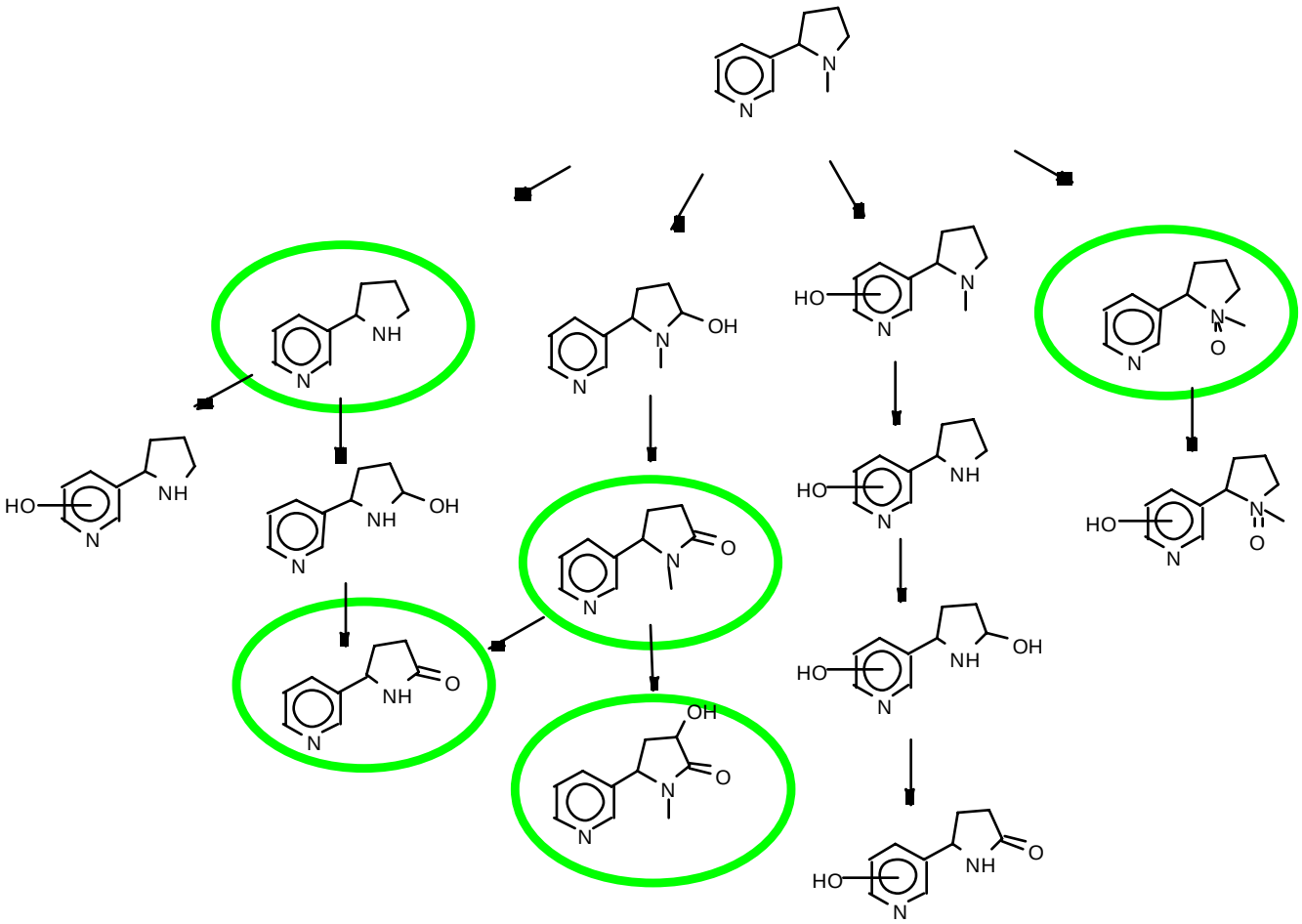


Developing Metabolic Simulators



Virtual metabolism uses a heuristic substructure search engine applied to a hierarchy of possible molecular transformations

Nicotine Simulated Metabolism

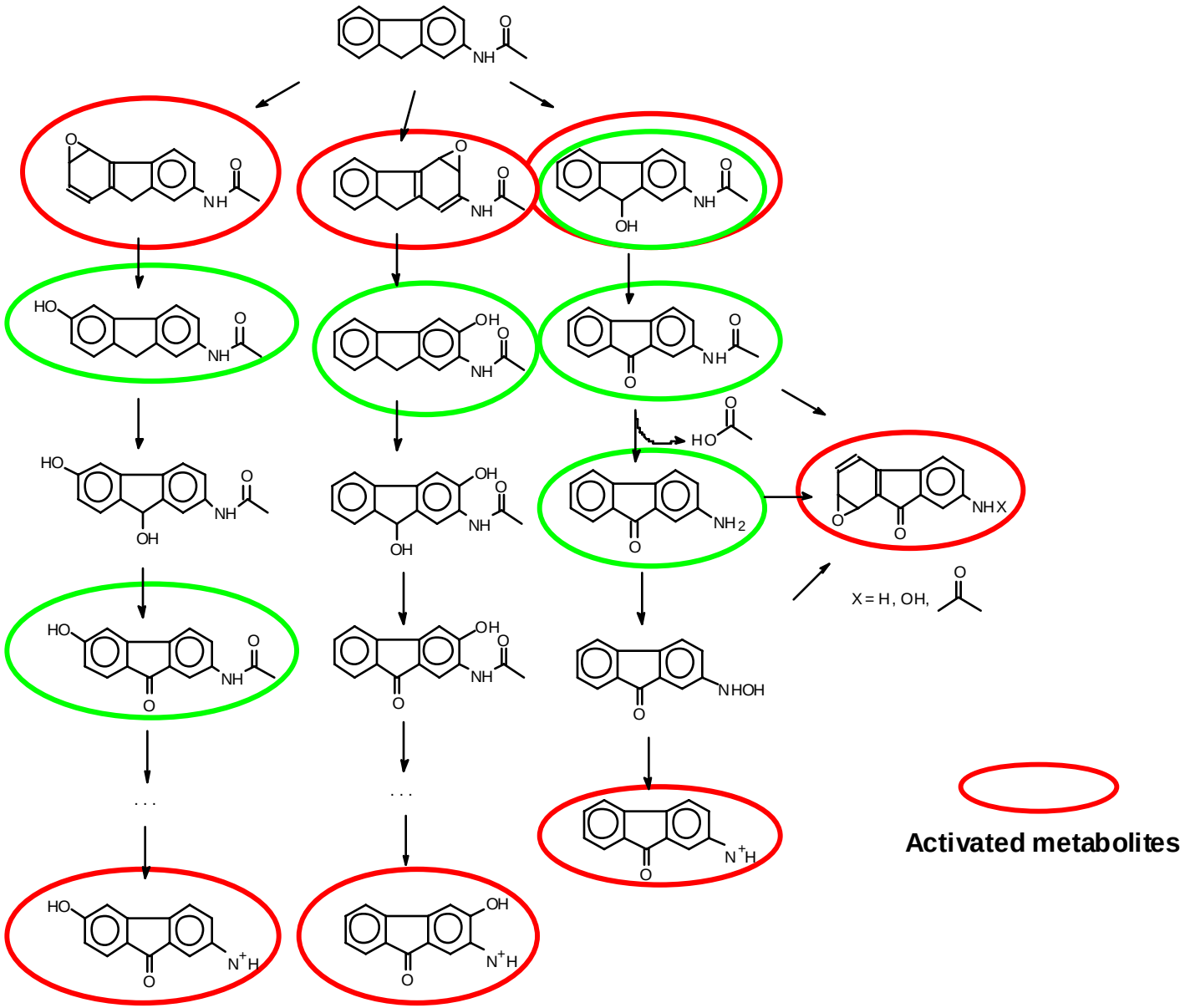


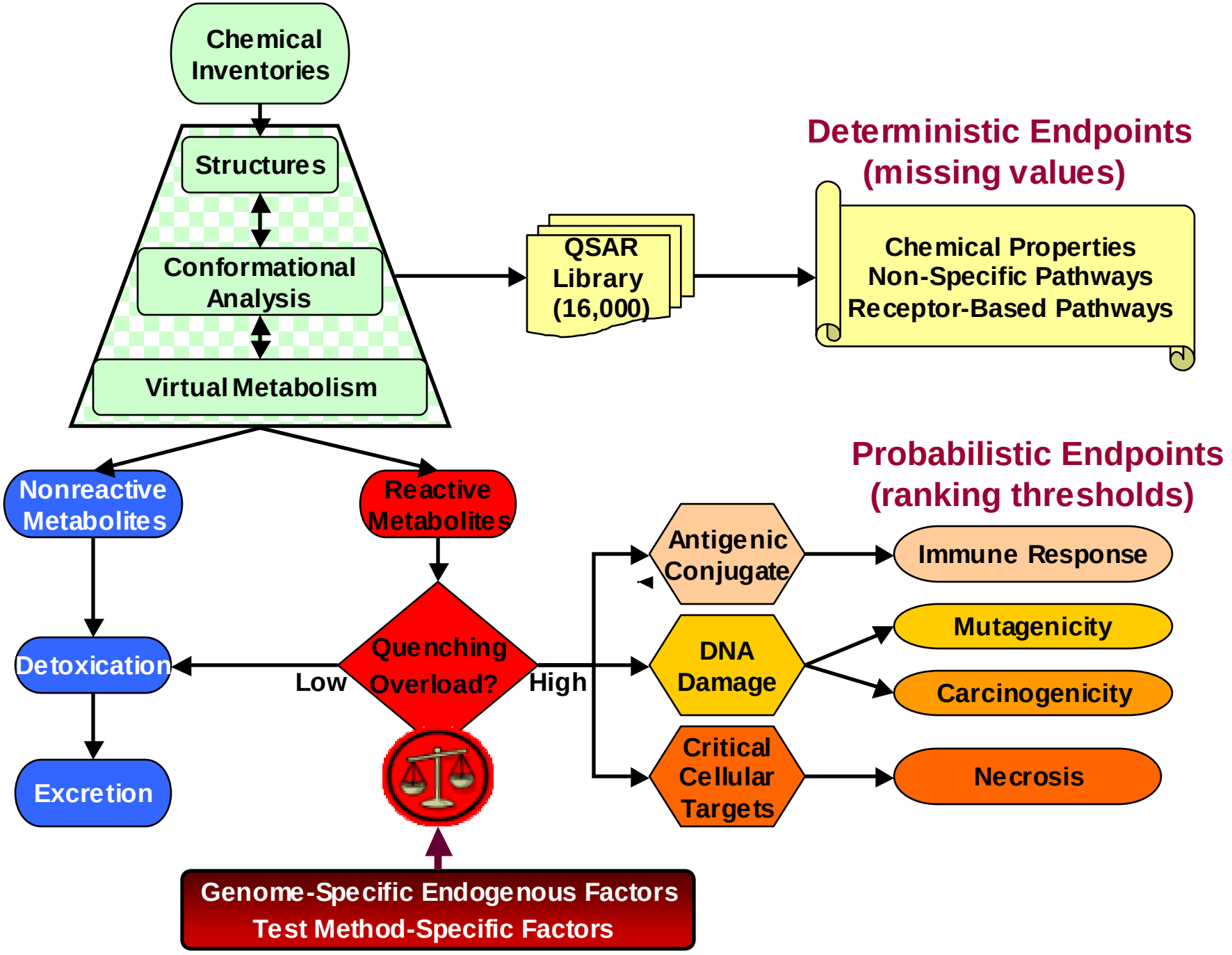
Confirmed metabolites in literature

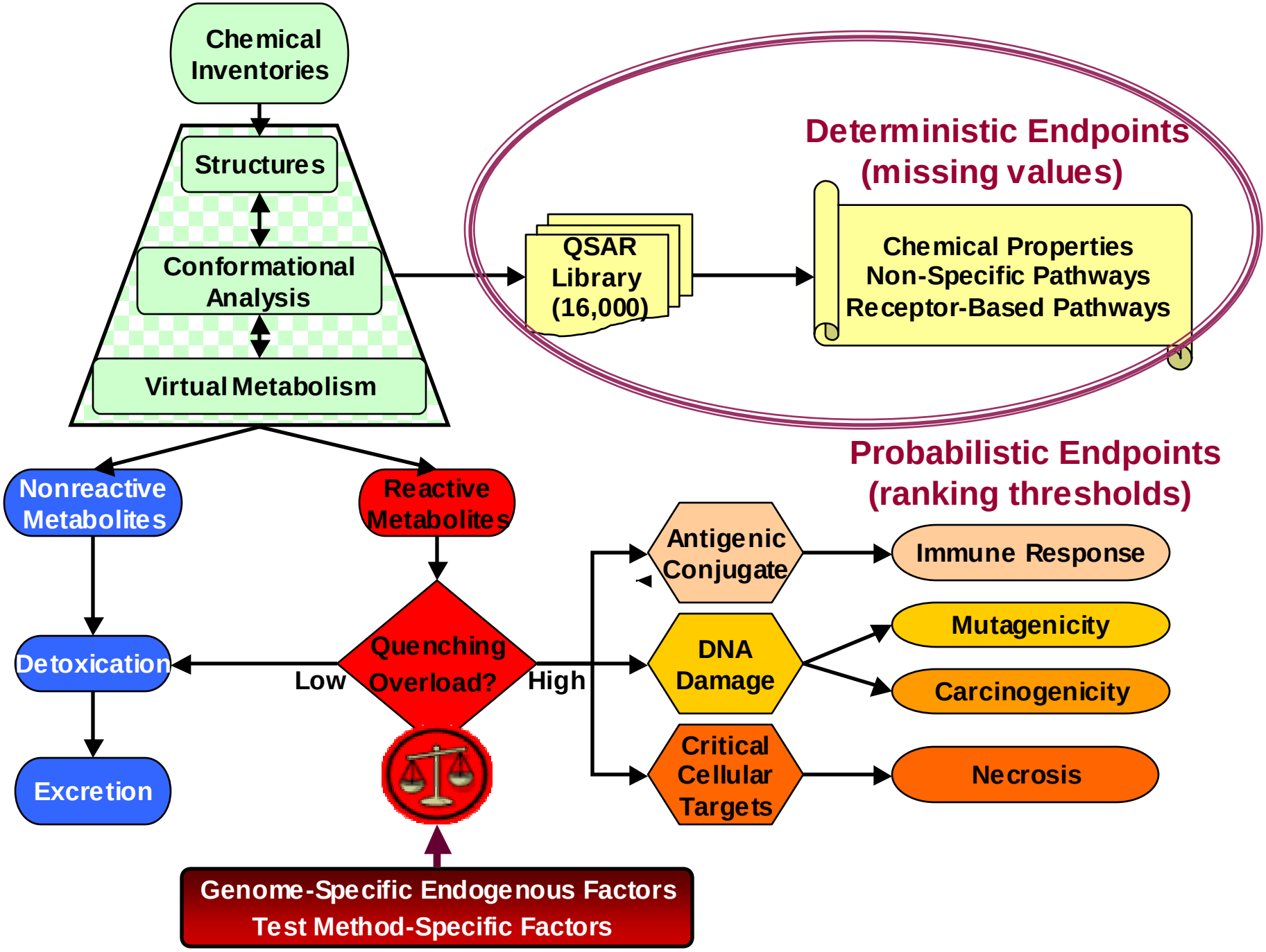
RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

2-Acetylaminofluorene Simulated Metabolism

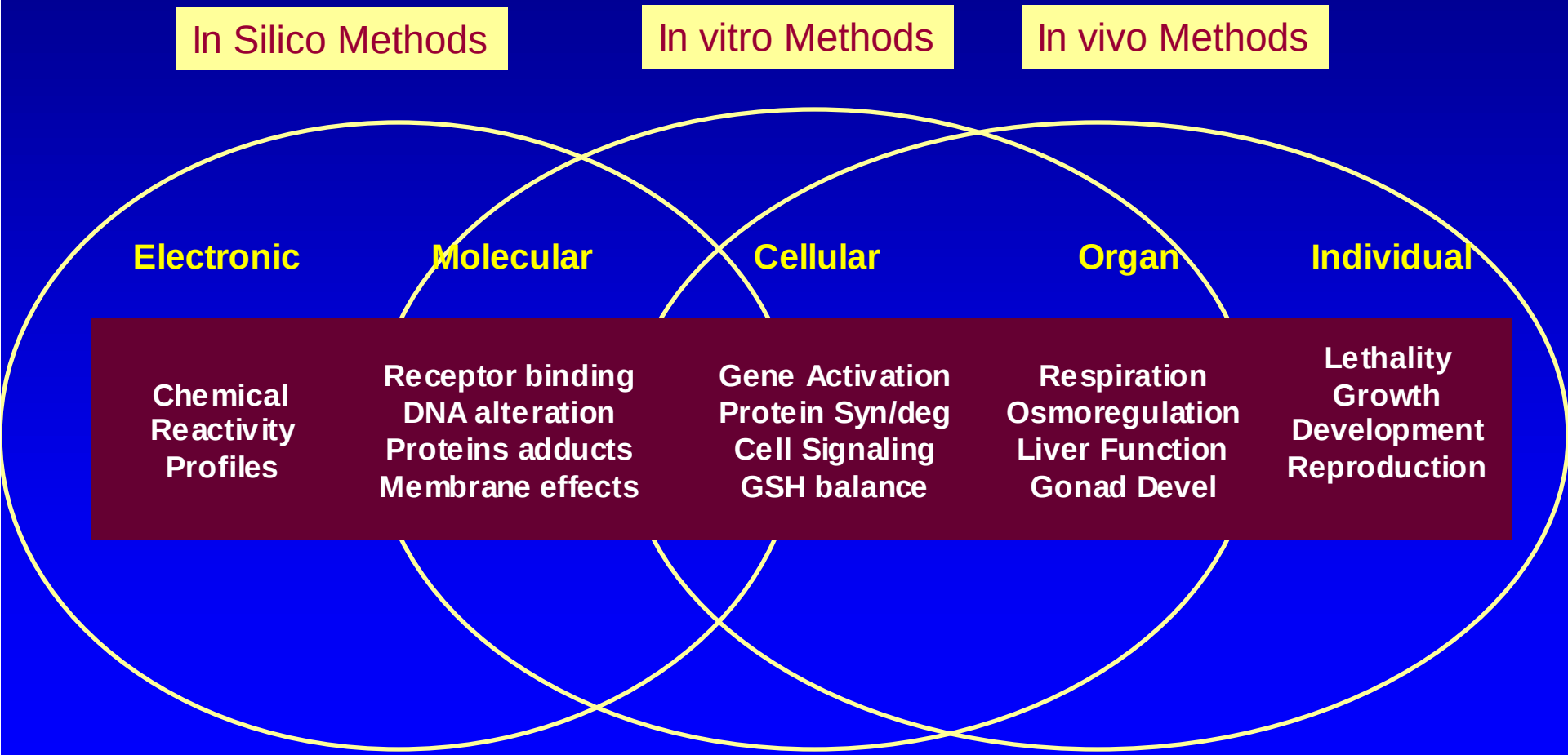



Chemical Inventories
Structures
Conformational Analysis
Virtual Metabolism
QSAR Library (16,000)
Deterministic Endpoints (missing values)
Chemical Properties
Non-Specific Pathways
Receptor-Based Pathways
Nonreactive Metabolites
Reactive Metabolites
Probabilistic Endpoints (ranking thresholds)
Detoxication
Excretion
Low
Quenching Overload?
High
Antigenic Conjugate
DNA Damage
Critical Cellular Targets
Immune Response
Mutagenicity
Carcinogenicity
Necrosis
Genome-Specific Endogenous Factors
Test Method-Specific Factors

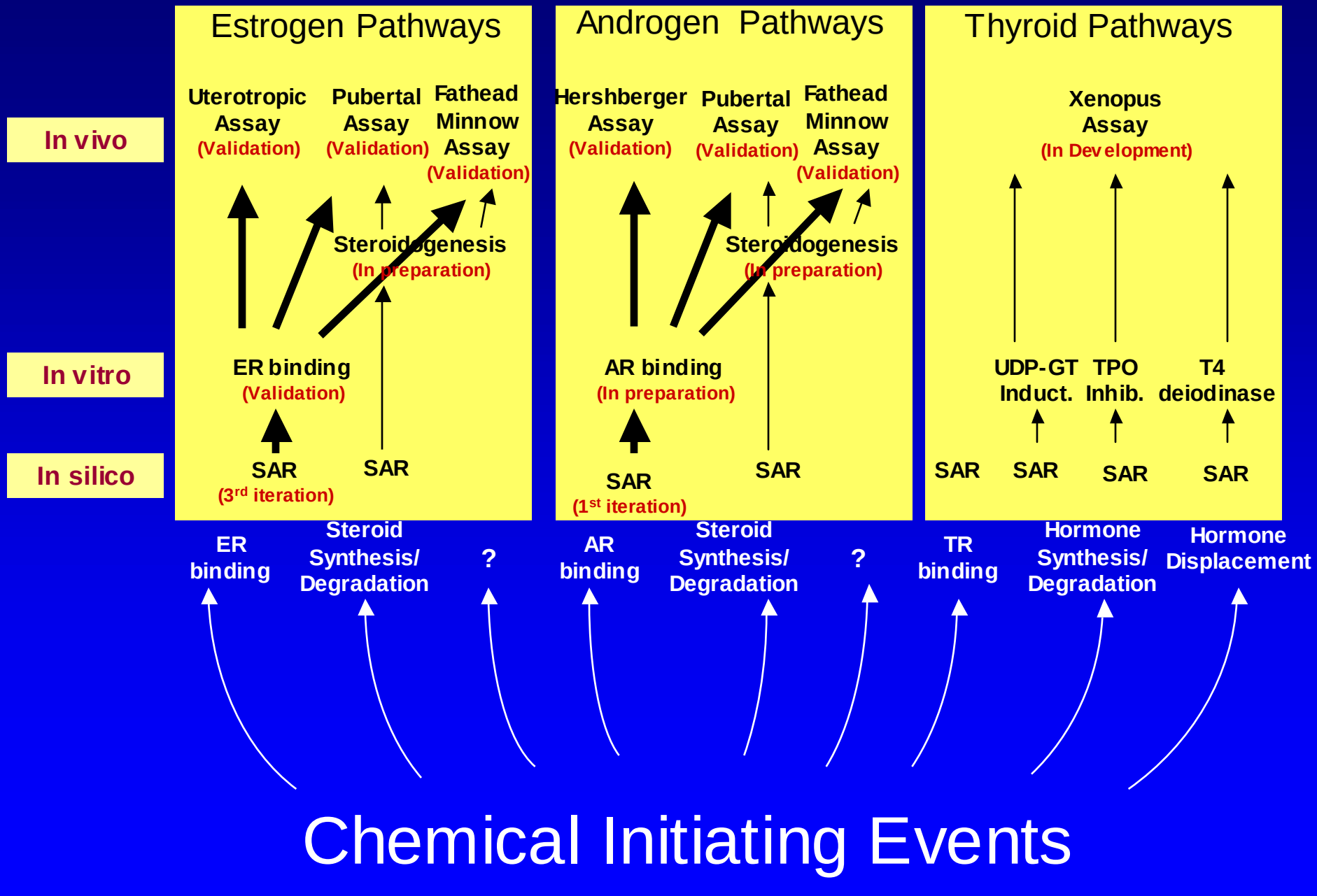

Chemical Inventories
Structures
Conformational Analysis
Virtual Metabolism
Deterministic Endpoints (missing values)
QSAR Library (16,000)
Chemical Properties
Non-Specific Pathways
Receptor-Based Pathways
Probabilistic Endpoints (ranking thresholds)
Nonreactive Metabolites
Reactive Metabolites
Detoxication
Excretion
Quenching Overload?
Low
High
Antigenic Conjugate
DNA Damage
Critical Cellular Targets
Immune Response
Mutagenicity
Carcinogenicity
Necrosis
Genome-Specific Endogenous Factors
Test Method-Specific Factors

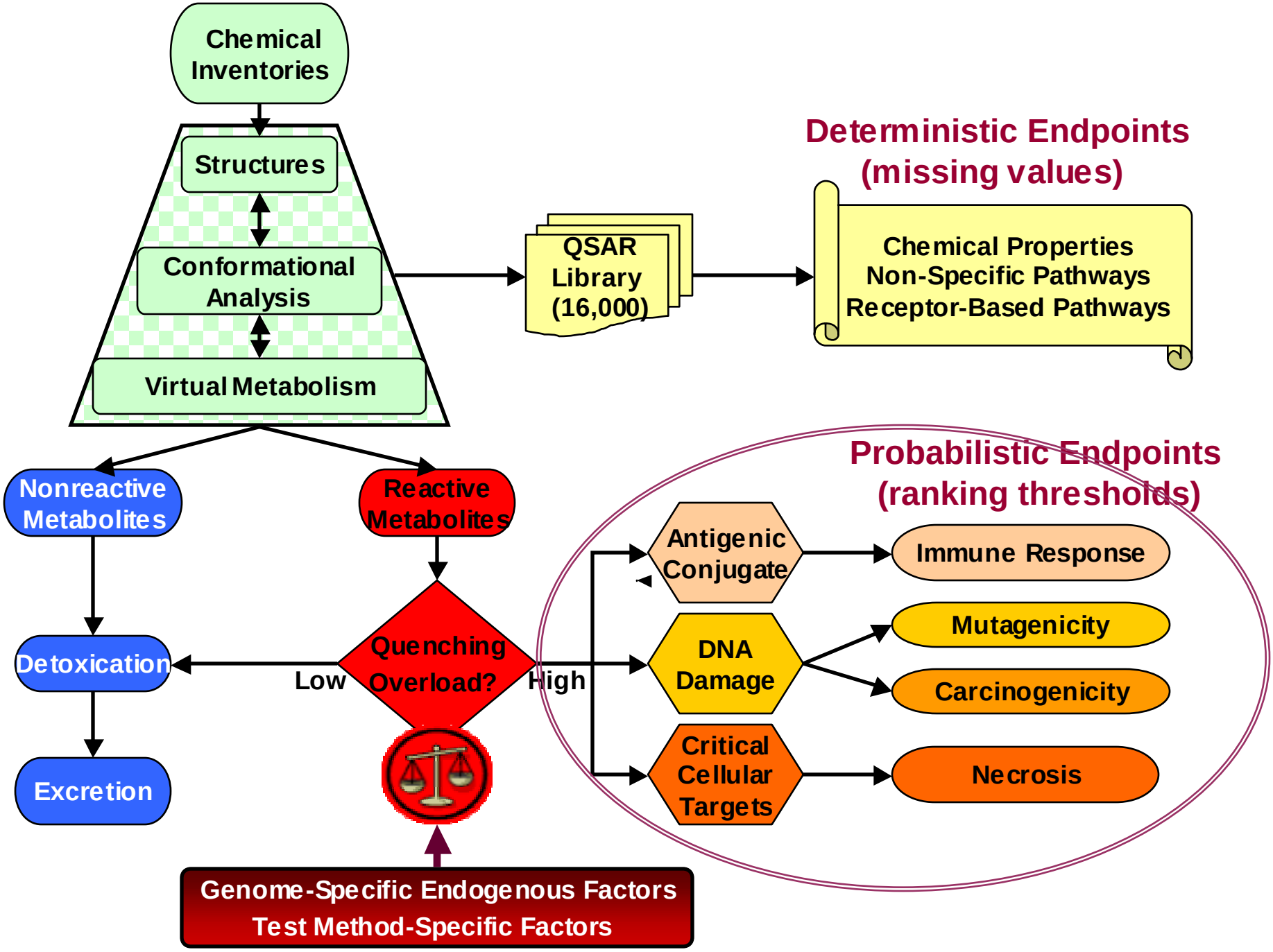
Delineation of Toxicity Pathways

Linkages Across Levels of Biological Organization



Impaired Reproduction/Development




Chemical Inventories
Structures
Conformational Analysis
Virtual Metabolism
QSAR Library (16,000)
Deterministic Endpoints (missing values)
Chemical Properties
Non-Specific Pathways
Receptor-Based Pathways
Nonreactive Metabolites
Reactive Metabolites
Detoxication
Excretion
Low
Quenching Overload?
High
Probabilistic Endpoints (ranking thresholds)
Antigenic Conjugate
Immune Response
DNA Damage
Mutagenicity
Carcinogenicity
Critical Cellular Targets
Necrosis
Genome-Specific Endogenous Factors
Test Method-Specific Factors

Probabilistic Models

$$P_{effect} = P_1 \times P_2 \times P_3 \times P_4 \times \ldots P_n$$

Forecasting distinct probabilities of low incident outcomes like idiosyncratic hepatic failure requires probability distributions for critical steps rather than effects under standard conditions

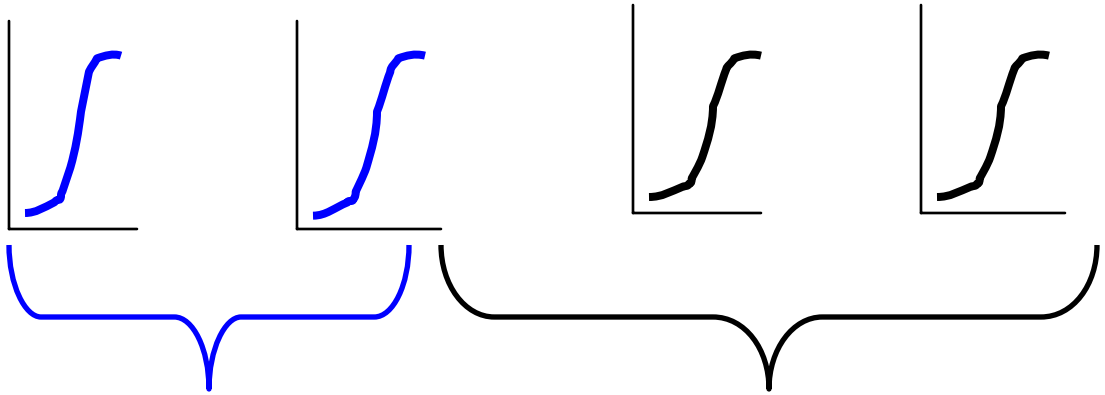
Exposure of the individual
Delivery rate to liver
Formation of reactive metabolites
Exceed detoxification rates
Covalent binding with proteins
Formation of neoantigens
Immune system recognition
Formation of cytotoxic antibodies
Interaction with hepatocytes
Overwhelm repair mechanisms
Liver Function Impairment
Liver Failure

--after Li (2002)

Probabilistic Models in Ranking

Prioritization does not require explicit QSAR estimates of toxicity for all chemicals but rather a reliable ordering of chemical activity with respect to explicit risk management scenarios

$$P_{effect} = P_{chem} \times P_{exposure} \times P_{environ} \times P_{genetic}$$



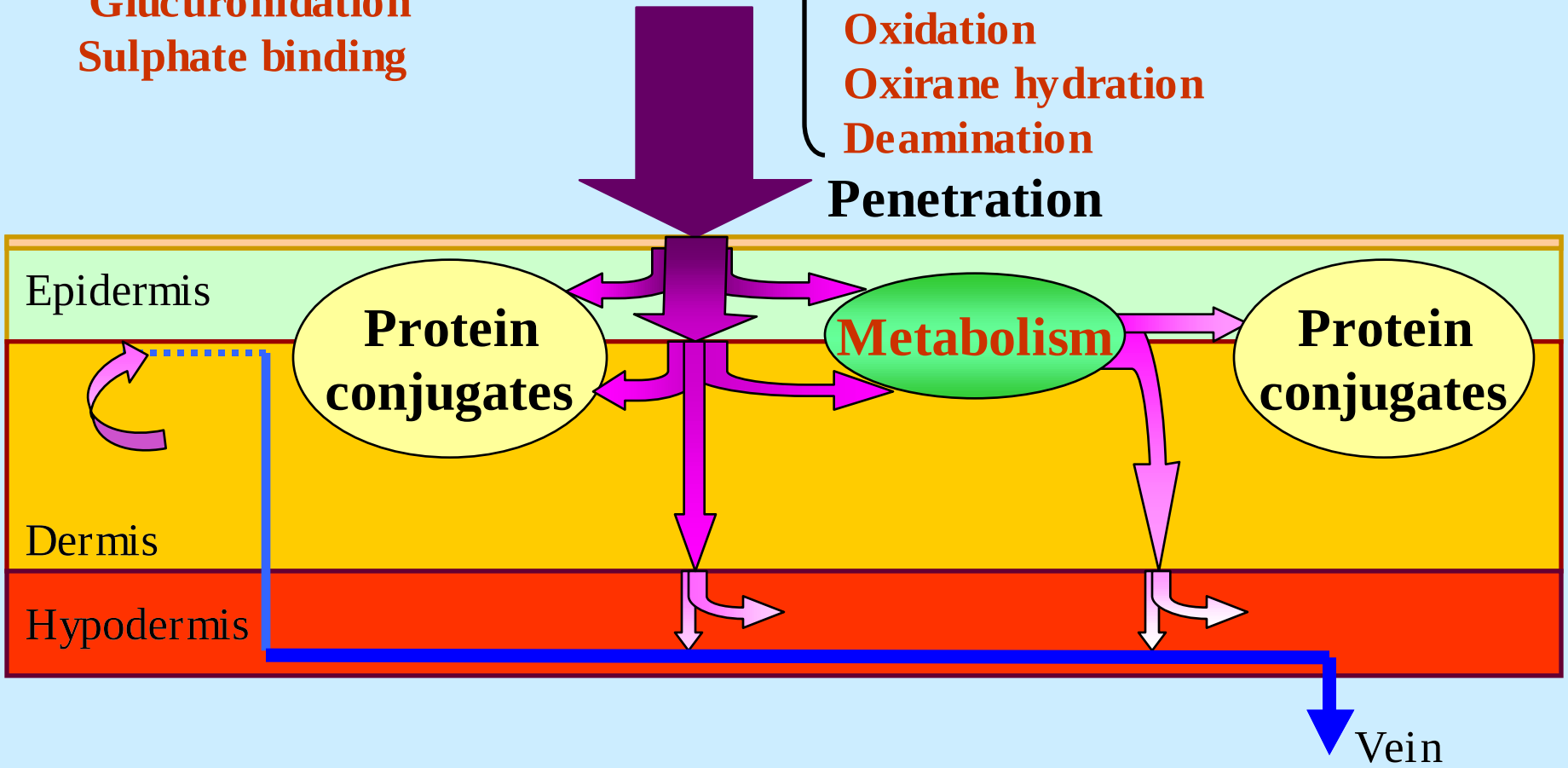
Skin metabolism

9 Spontaneous reactions
85 Phase I reactions
12 Phase II reactions

Decompositions of geminal derivatives
Decompositions of hydroperoxides
Keto-enol tautomerism

C-Hydroxylation
Epoxidation
Ester hydrolysis
Dealkylation
Oxidation
Oxirane hydration
Deamination

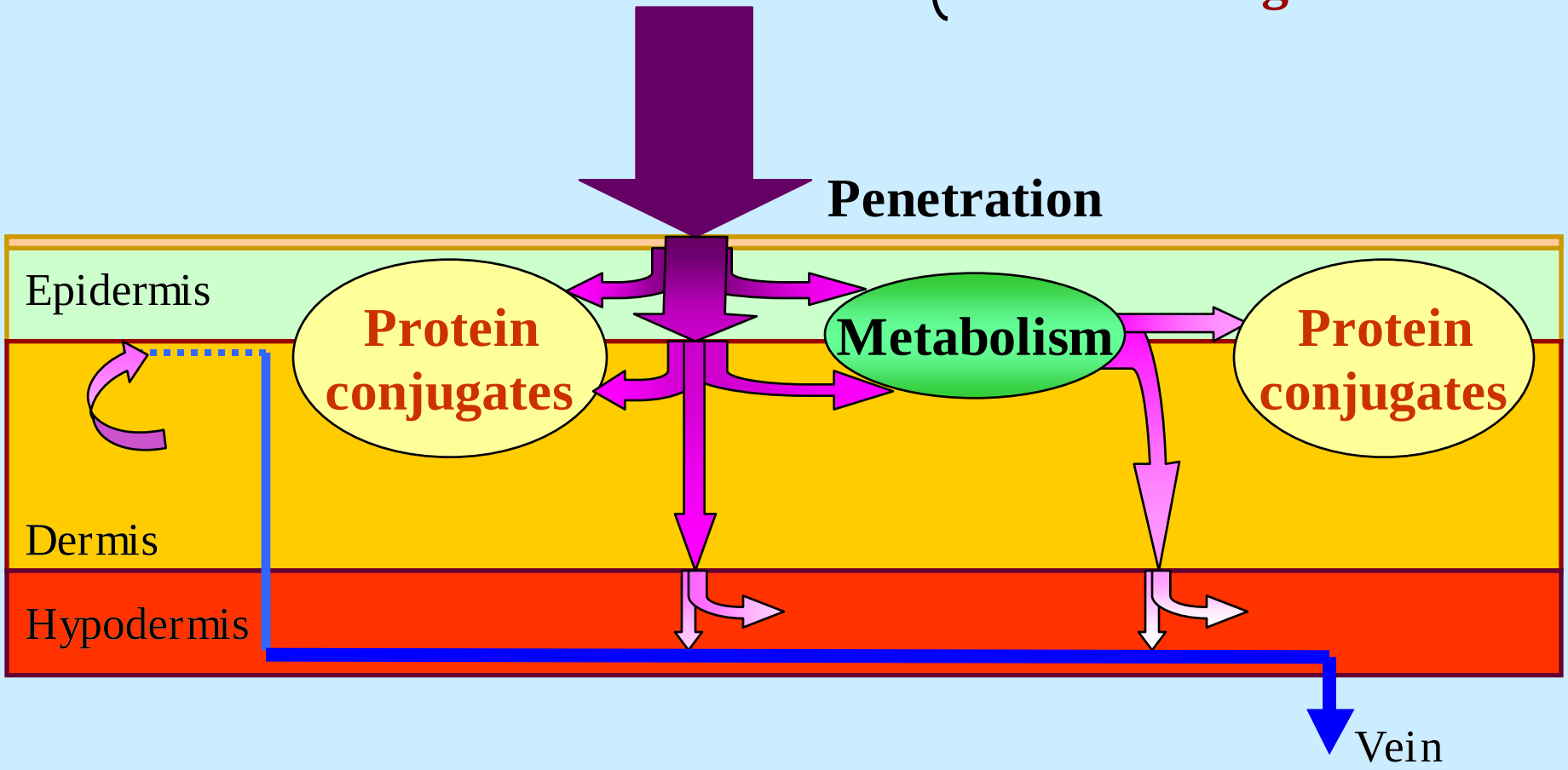
Glucuronidation
Sulphate binding



Reactions with skin proteins

~50 reactions of protein including

- Nucleophilic substitution
- Schiff-base formation
- Michaels addition
- Alkylating cyclic agents
- Free radical generation



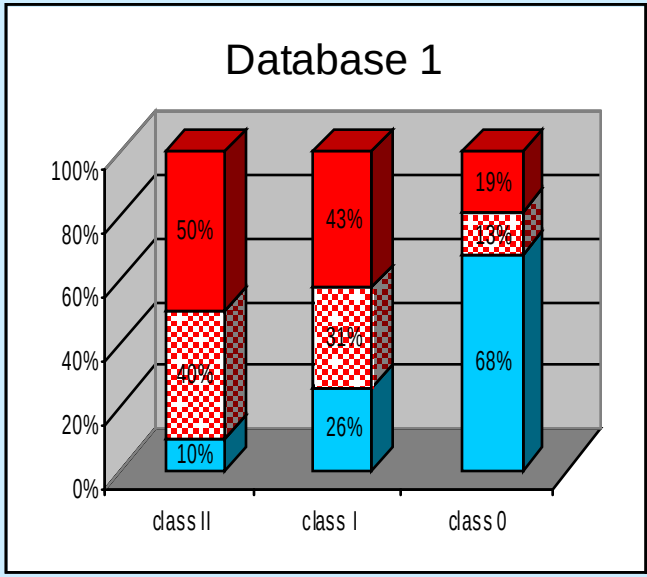
Skin Sensitization Databases

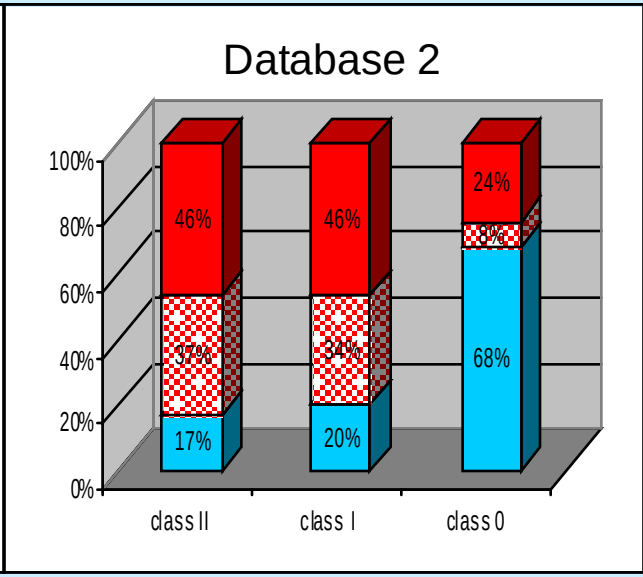
Category	Database 1 249 chemicals	Database 2 192 chemicals		Database 3 307 chemicals
		Human allergen	EC3 range	
Class II	Strong sensitizers 103 chemicals	Strong Moderate 92 chemicals	< 0.1% 0.1-10%	Strong sensitizers Moderate sensitizers 161 chemicals
Class I	Moderate sensitizers 77 chemicals	Weak 41 chemicals	10-30%	Weak sensitizers 55 chemicals
Class 0	Weak sensitizers 69 chemicals	Extremely weak Non-sensitizer 59 chemicals	30-50% >50%	Non sensitizers 91 chemicals

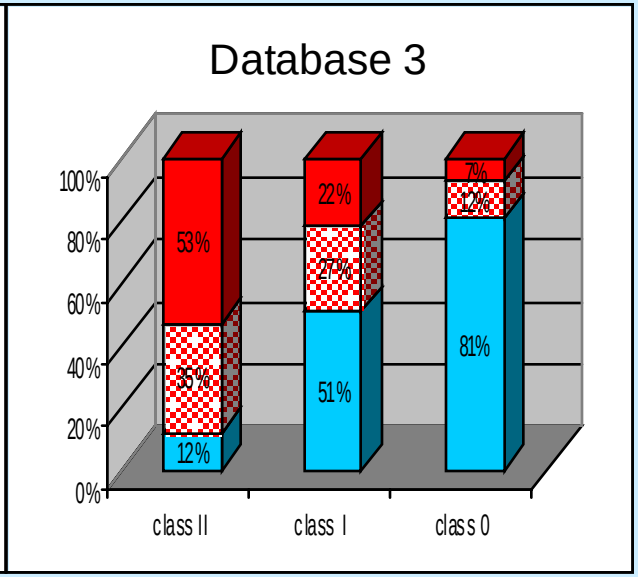
Class II
Class I
Class 0

1st Iteration QSAR-based Model for Sensitizers

- **Parent chemical reacts directly with protein(s)**
- **Predicted metabolite(s) react with protein(s)**
- **No reaction with proteins predicted**







Complexity in Quantifying Chemical Structure

- Conformational Analysis

  Chemical properties/interactions are highly sensitive to induced conformational changes

  - which structure should be quantified?

- Metabolic Simulation

  Many toxic effects do not arise from nor can they be forecasted directly from the parent chemical
  - which metabolite should be modeled?

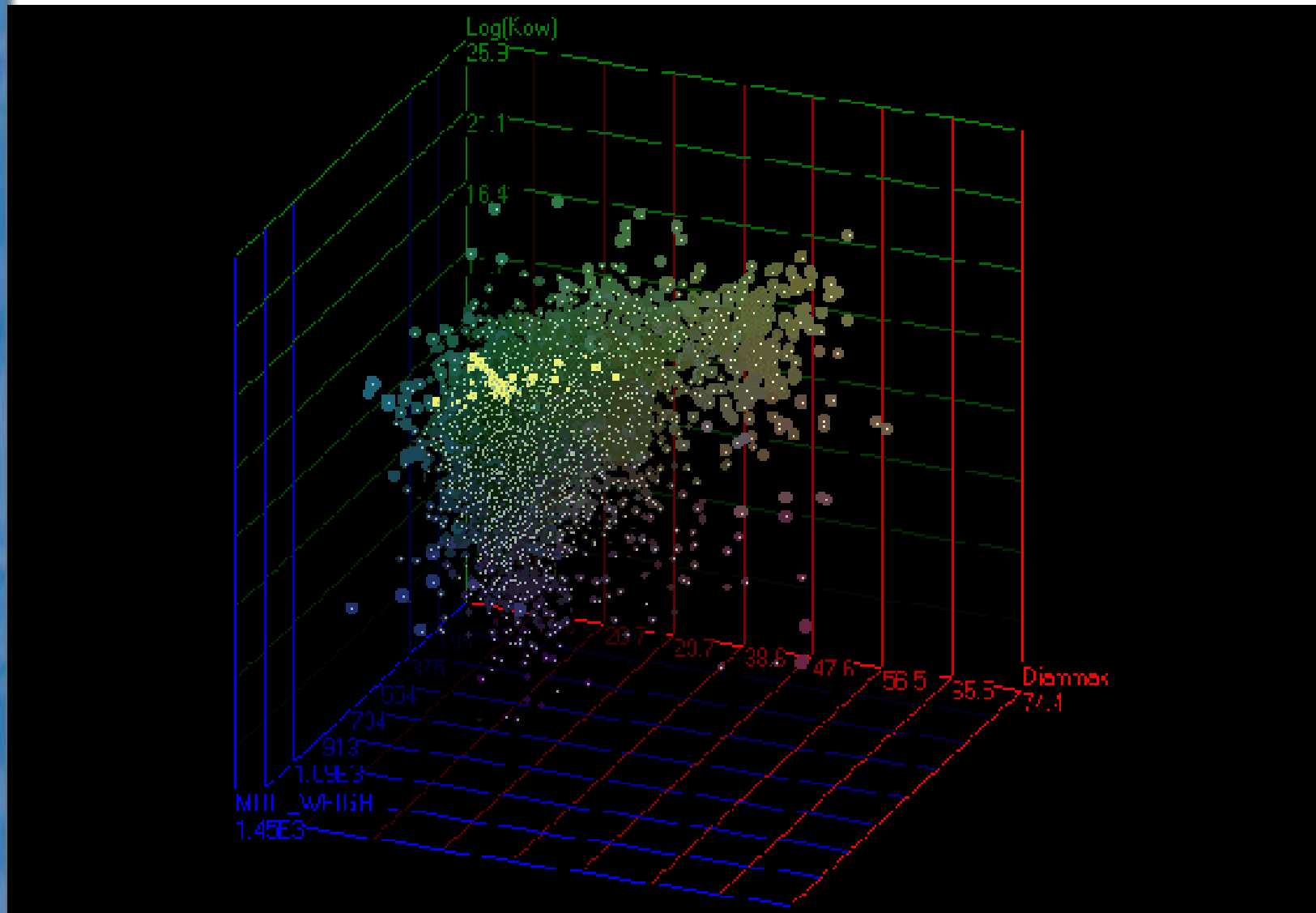
- **Chemical Inventories and Scaffolds**

  Environmental risks include those from the few tested chemicals and thousands of untested chemicals
  - which group/list should be EPA's priority?

RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

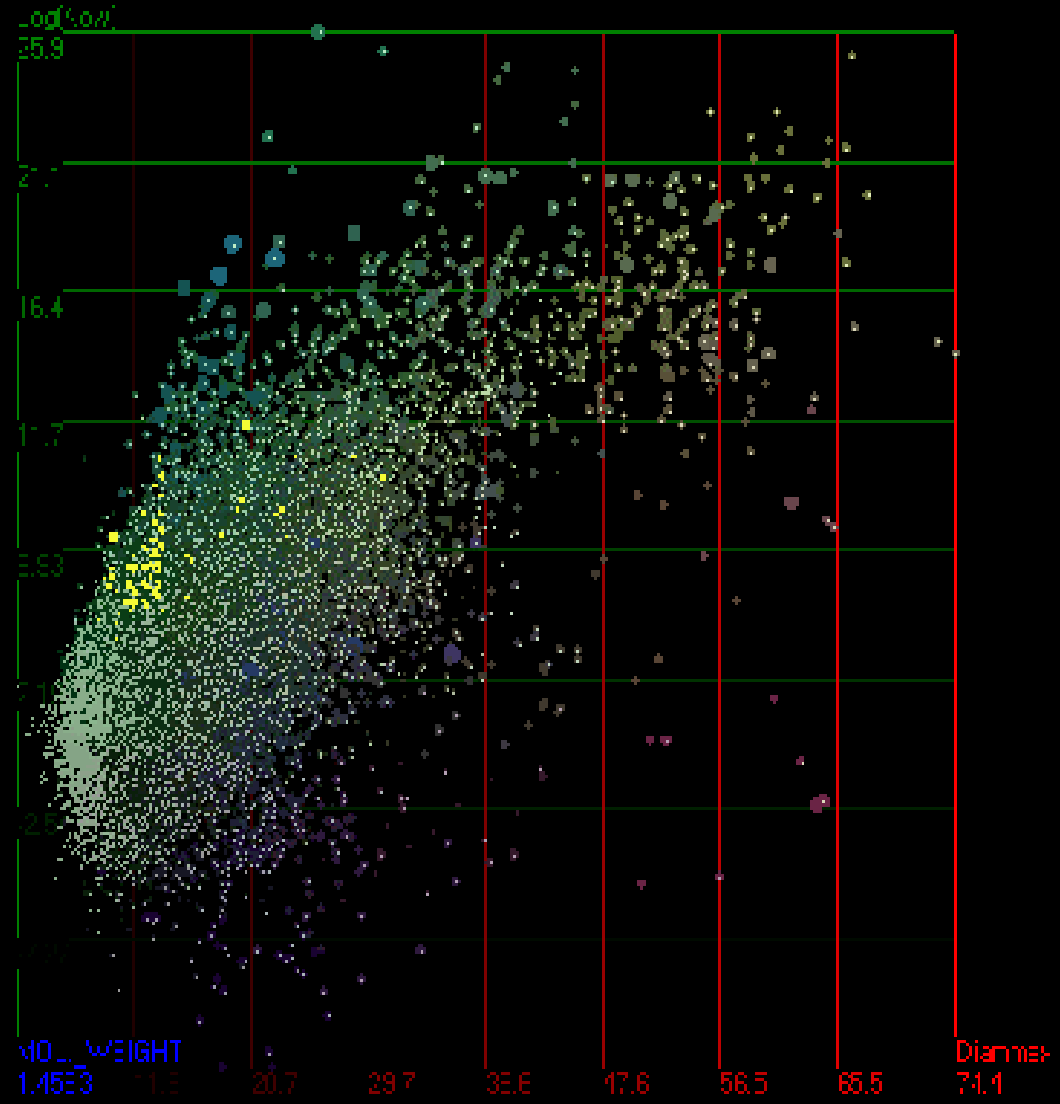
Structural Domain of Found Data[1]



RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Structural Domain of Found Data2

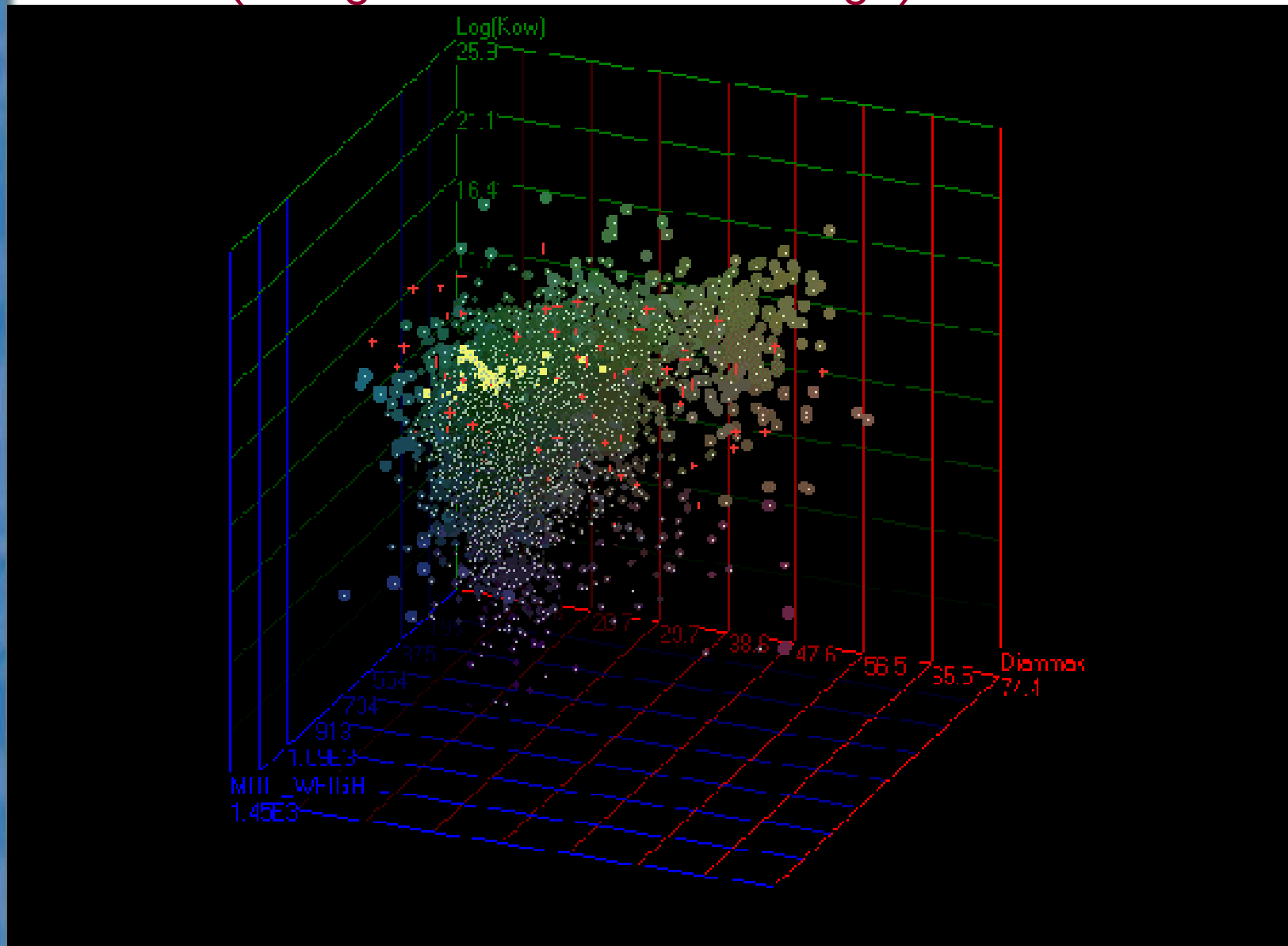


RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Expanding Knowledge-Bases Strategically

(Design for Uniform Coverage)

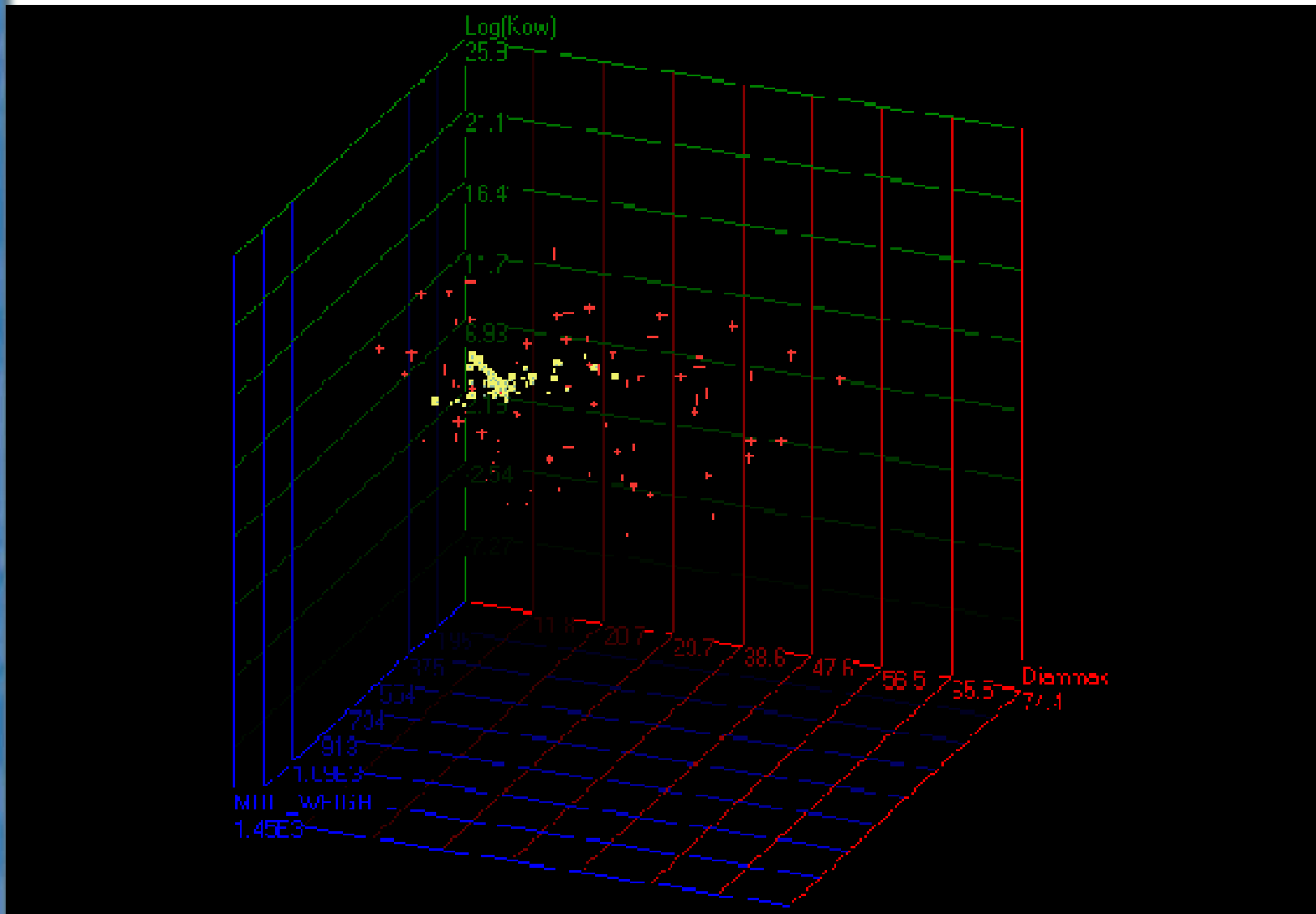


RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Strategic Expansion of Knowledge-Bases

(Design for Uniform Coverage)

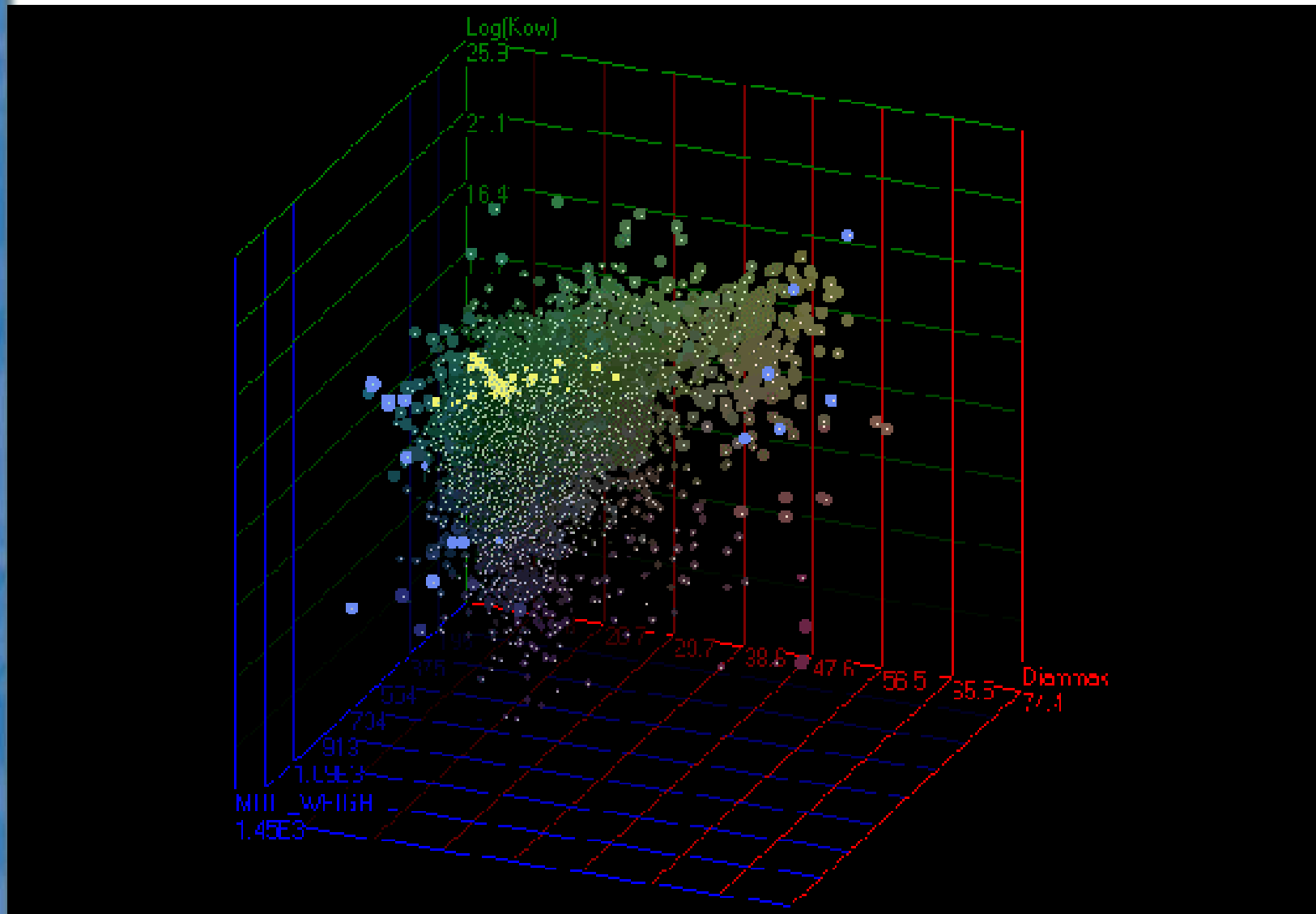


RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Strategic Expansion of Knowledge-Bases

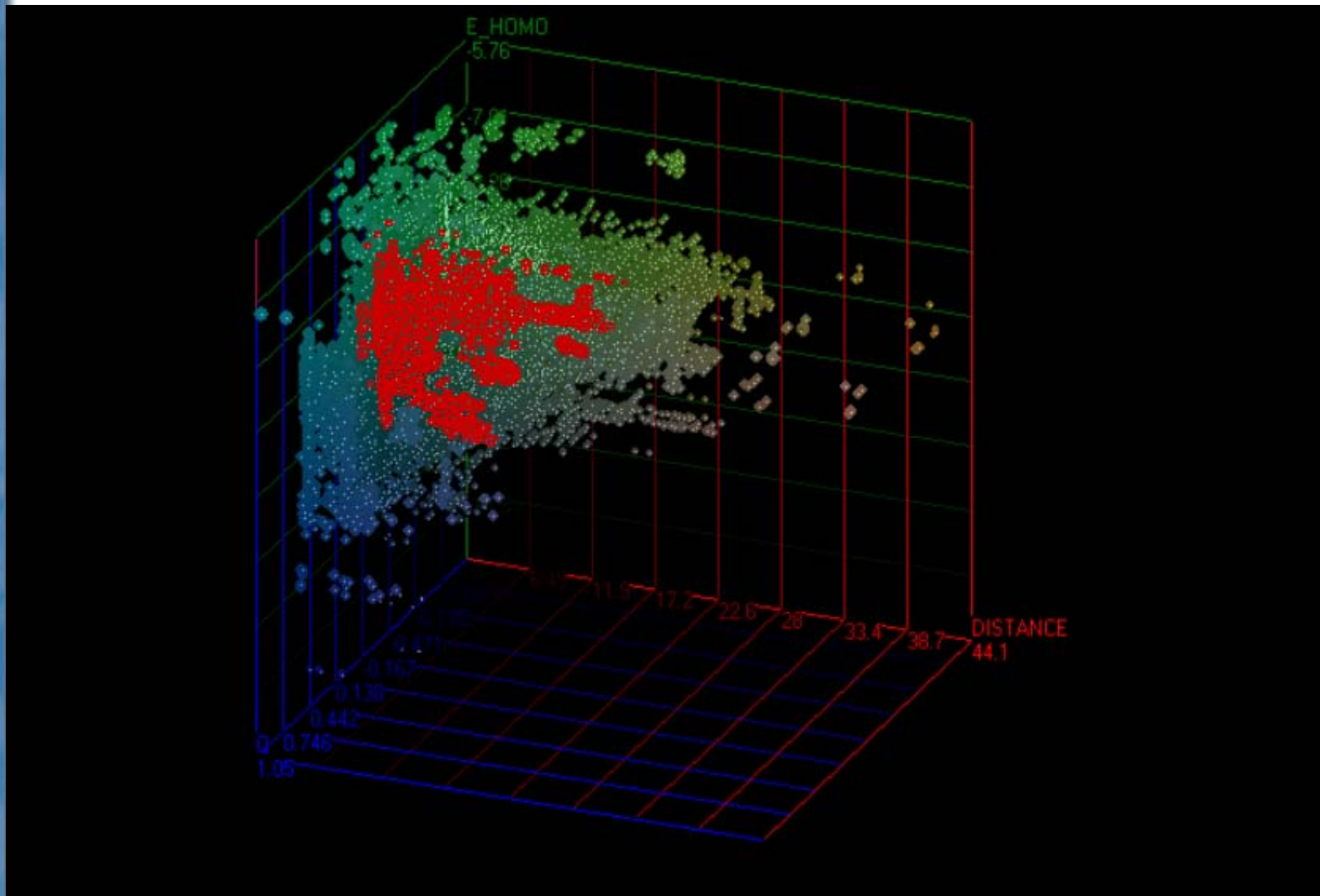
(Design for Structural Extremes)



RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

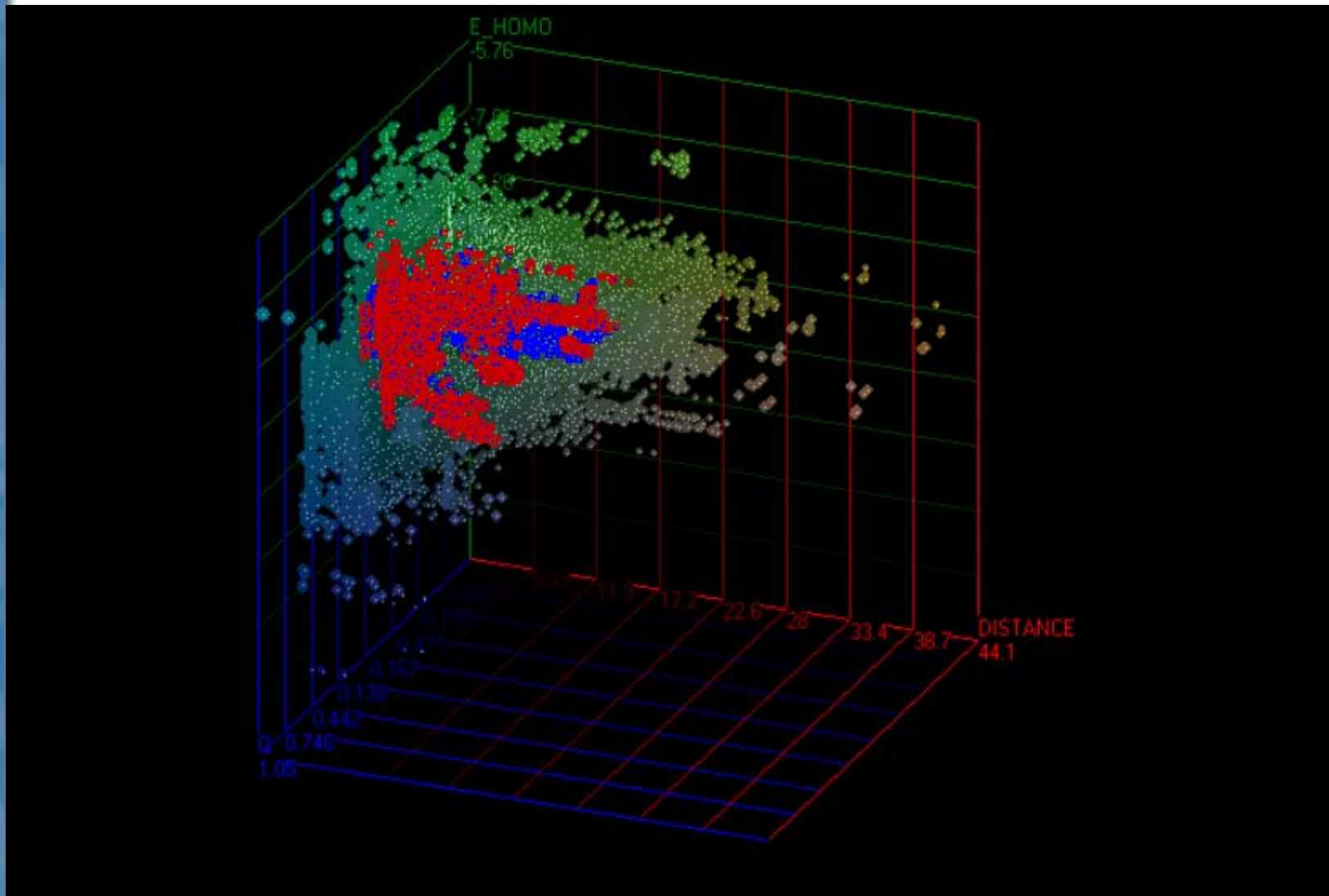
ER_{rat}-Binding Affinity Knowledge Base
(Found Data in Red)



RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

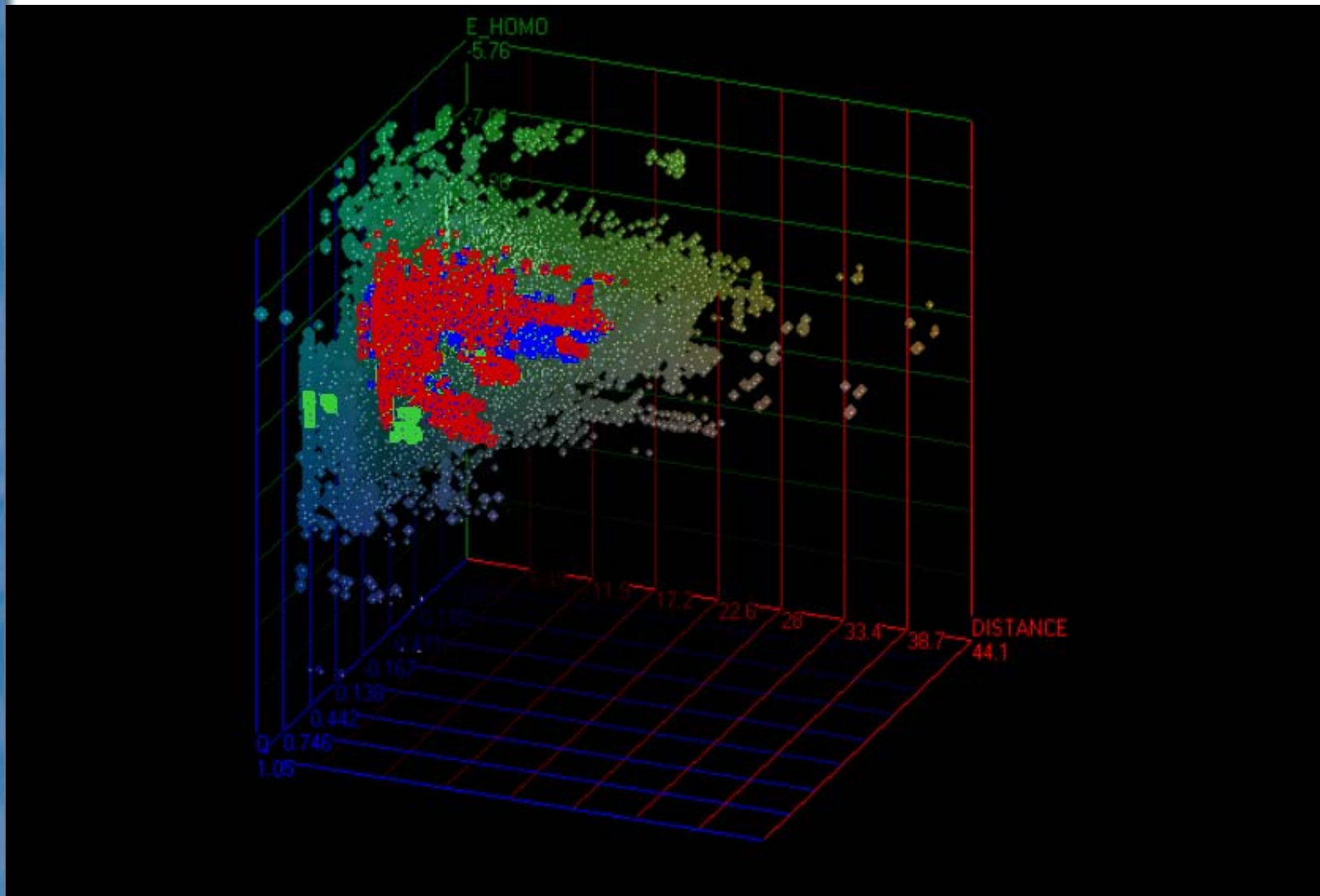
$ER_{rat/mouse}$-Binding Affinity Knowledge Base



RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

ER-Binding Affinity Knowledge Bases
(rat, mouse,human)



-QSAR-
Experimental Design or Luck-of-the-Draw

- *Post hoc* QSAR modeling of "found" data proliferates localized models which become spurious in a larger context
  - localized models cause major errors in global prioritizations
- Global models emerge from strategic chemical selection in a hypothesis-testing-hypothesis testing process with important endpoints
  - 20 well-chosen chemicals can give more mechanistic information than 500 exploratory tests (ER example)
- Defining the chemical boundaries (domain) of a screening model is often more important than improving the model statistics
  - EU is leading effort to develop regulatory acceptancy criteria

Computational Toxicology Challenges

- Create a virtual escape from the EDC screening dilemma for EPA chemical lists and inventories
  - *in silico* identification of chemicals for further laboratory testing
  - balance exposure-toxicity risks in screening large inventories
- Develop a Systems Biology approach that links test-specific effects as localized symptoms of more global molecular events
  - Redefinine chemical reactivity in a hazard identification context
  - add “likelihood estimates” to the spectrum of possible effects
- Provide a scientific foundation for a hypothesis-driven testing paradigm for EPA risk assessment processes
  - Introduce risk management thresholds along toxicity pathways
  - reduce animal testing by minimizing “negative” laboratory tests